

## ***Fitting Your Theory to the Facts: Probably Not Such a Bad Thing After All***

### **1. Introduction**

In the following pages I shall try to show that the variety of Bayesian confirmation theory based on so-called personal probabilities provides an intuitively correct solution to an outstanding problem in a controversial area of methodology, and that it does so in an entirely natural and unforced way. In the course of the discussion it will, I hope, become apparent that the objections usually thought to be decisive against Bayesian theories in general, and especially against this one, are in reality nowhere near as damaging as they seem; and I shall end by making the possibly surprising claim that Personalist Bayesianism offers just the sort of theory of confirmation, fallibilistic in temper, founded on deductive principles only, yet fruitful in methodological information, with which even Popper himself should not be able to find fault (though he would, and does, of course, because it is 'subjective'; but more on that later).

### **2. The Null-Support Thesis**

The outstanding problem is the ancient one of what epistemological distinction if any we are entitled to draw between a theory that has independently predicted an observed effect and one that has been deliberately constructed to yield the effect as a consequence. One answer to that question, and an extremely popular one, is to deny that the second theory derives any support at all from its 'prediction' of the effect in question. An apparently powerful argument in support of this position is the following. If we were to concede that the second theory is supported by the known effect, we should, or so it appears, be faced with the awkward, if not intolerable, consequence that it becomes a simple matter to generate arbitrarily many theories that are supported by any given piece of data, but that are, intuitively speaking, not really supported at all. For example, suppose that A's heuristic deliberations result in a hypothesis, call it H, that x and y stand in a specific functional relation  $y = f(x)$ , and that A then tests H with a large number

$n$  of joint observations  $(x_1, y_1), \dots, (x_n, y_n)$  of  $x$  and  $y$ . Suppose also that, within the given error bounds, the observations all lie on that curve. While we will probably agree that the observations confirm  $H$ , we should equally probably be reluctant to concede that they also support each of the infinitely many hypotheses determined by a particular choice of  $g$ :

$$y = f(x) + (x - x_1)(x - x_2) \dots (x - x_n)g(x) \quad (1)$$

(These ad hoc variants, essentially precursors of Goodman's 'grue' hypothesis, were introduced into the literature by Jeffreys [1948] 3).

Non-Bayesian theories have difficulty with hypotheses like (1); that the observations do not by themselves appear to discriminate between them and  $H$  is after all precisely the rub of Goodman's paradox. Bayesian theories at least have the formal capacity to discriminate by means of an appropriate distribution of prior probabilities: to what extent this is successful is a question I shall answer later. Without leave to appeal to considerations of plausibility prior to the data (possibly, as in Jeffreys's Bayesian theory, which I shall discuss briefly later, based on considerations of simplicity, in some one of the numerous explicata of that troublesome notion), there seems to be no intrinsic difference between  $H$  and its variants (1) that justifies denying them all equal status in explaining and being supported by the observations. But there is an extrinsic, but apparently quite objective, difference: the hypotheses (1) are constructed specifically to fit the data, while  $H$  was not. According to what I shall henceforward call the null-support thesis, the variants (1) are not supported by the  $n$  observations, precisely because they were deliberately contrived to explain those observations.

The null-support thesis has a long and respectable pedigree: it is to be found in Bacon, Descartes, and Leibniz; it is a principle incorporated into Popper's theory of corroboration, and we find it recently endorsed by Zahar (1989, 16), Redhead (1986), Giere (1984, 159–61), and Worrall (1978, 48). In what follows I hope to show that, despite this powerful advocacy, the null-support thesis is false. Indeed, I shall produce some hypotheses that well-established canons of scientific procedure pronounce very strongly supported by the data they were constructed to explain. It follows that the reason we tend to depreciate the ad hoc alternatives (1) cannot be that they are constructed from the data. What is it then? It is, I shall argue, just because they are irremediably ad hoc; because, in other words, the structural model implicit in the parametric hypothesis: there are  $a_1, \dots, a_n$  such that for all  $x$ ,

$$y = f(x) + (x - a_1) \dots (x - a_n)g(x),$$

where  $g$  is not identically zero, is thought not to correspond to the actual state of affairs; we just don't think it true. And in general, hypotheses simply contrived to fit the data will be looked askance at, not because they have been made to fit the data, but because they have been contrived *simply* to fit the data; which is just

another way of saying that we have, as far as we can tell, no independent reason to believe them true.

These remarks are not as empty of explanatory value as they might appear. The falsity of the null-support thesis means that a more flexible account of support is required, which, in the course of explaining when and for what reasons we ascribe support to hypotheses generally, explains also why we discriminate *within* the class of hypotheses constrained to fit specific pieces of data, as to which merit support therefrom and which do not. I shall provide such an account. It is, I shall argue, implicit in the Personalist Bayesian theory, which does indeed say that the variants (1) will be assigned vanishingly low support if the structural models implicit in them are sufficiently strongly disbelieved—as indeed, by assumption, they are. But the most important consequence of the Bayesian theory is the fundamental principle of all inductive inference: evidence supports a hypothesis *h* the more, the less it is explicable by any plausible alternative compared with its explicability by *h*. This is the criterion by which in practice we decide which data support a hypothesis and which do not, and it is a criterion that can just as easily be satisfied by a hypothesis constructed from that evidence as by one that was not.

Much of the motivation for the null-support thesis comes, I think, from a desire to legitimate a preference, which it is alleged that a study of the history of science reveals, for hypotheses that independently predict facts over those that merely accommodate them. Whatever else this preference reflects, however, it is not that the former are always better supported by those facts than the latter. For there are convincing counterexamples to that doctrine too. Lest it should be thought that things are now going too far in a direction away from the null-support thesis, I shall show that the Bayesian account explains why, of two rival theories, initially equally well supported, but differing in that one independently predicts data that the other merely absorbs into the evaluation of a free parameter, the former receives the greater support from those data. This result is, I think, the germ of truth in the generally false thesis that the independent prediction of facts invariably merits greater rewards of support than their post hoc explanation.

These conclusions will be shown to follow straightforwardly. No arbitrary assignments of prior probabilities, no fiddling or gerrymandering the formulas are necessary; what will emerge from the discussion is just how essentially Bayesian our informal reasoning is. I recognize that any such conclusion is bound to be greeted with strong reservations in many quarters, and I shall conclude this paper by trying to show that the usual objections advanced against allowing the Bayesian theory a role in 'objective' methodology are unsound.

Some preliminary words are in order about the Bayesian theory that is being credited with these virtues. That theory is, as I have already said, the one that, following L. J. Savage, has come to be called Personalist Bayesianism. It exploits the fact, explicitly stated and proved only in this century, but taken for granted from the outset, that the probability calculus furnishes the fundamental laws of

fair odds, or to be more precise, fair betting quotients (betting quotients  $p$  are related to odds  $x$  by the equation  $p = x/(1+x)$ , with inverse  $x = p/(1-p)$ ). Fair odds are odds that would give no advantage to either side of a bet were one to be called; whether any such odds exist, however, or what they are if they do, is, except in a very restricted class of cases, a matter on which the theory eschews an opinion.

The significance of setting out the general laws that any system of fair odds must obey is that from the very first, when people started talking about the probabilities of hypotheses, they habitually glossed those probabilities as what would be, relative to the contemporary state of knowledge, the fair betting quotients in the ideal situation that the bets could be unambiguously settled. Now a famous result, proved in this century independently by Frank Ramsey and Bruno de Finetti, implies that if any set of betting quotients fails to satisfy the calculus of probabilities, then those betting quotients cannot all be fair. What Ramsey and de Finetti in fact showed was that if an opponent is free to dictate which side of a bet you will take, and the stakes on each, then were you to engage in simultaneous bets with betting quotients that do not satisfy the probability calculus, you could be forced to make a loss (or gain) come what may. But if a betting quotient is fair, the advantage to taking a given side of a bet should be zero; and the net advantage relative to a system of bets at fair betting quotients should also be zero, since it is a sum of zeros. So if someone could tell in advance that a particular betting strategy would entail a positive loss or gain, and hence presumably that the net advantage at those odds cannot be zero, then it follows that the odds are not all fair. Ramsey's and de Finetti's theorem tells us, therefore, that the rules of the probability calculus are nothing more than consistency constraints on the construction of a set of subjectively fair betting quotients, i.e., betting quotients that you believe, in the light of your available information, determine fair odds.

The fundamental notion in the application of this theory to the problem of inductive inference is that of the conditional probability,  $P(h/e)$ , of  $h$  relative to  $e$ . Your conditional probability of  $h$  on  $e$  is what you would assess the probability of  $h$  as, were you to come to know  $e$  (but nothing more).  $P(h/e)$  is, by Bayes's theorem, equal to  $P(e/h)P(h)P(e)^{-1}$ .  $P(e/h)$  is called the likelihood of  $h$  on  $e$ , and is equal to one if  $h$  entails  $e$  (modulo initial conditions and the more or less extensive quantity of other background information you are equipped with). Where  $h$  describes a hypothetical physical probability distribution over a set of data points, some measurable subset of which is defined by  $e$ , then  $P(e/h)$  is set equal to the probability  $h$  ascribes to  $e$ .  $P(h)$  and  $P(e)$  are the so-called prior probabilities of  $h$  and  $e$  respectively.  $e$  is reckoned to support  $h$  if  $P(h/e) > P(h)$ , and the difference between the two is a useful measure of the extent of this support, which clearly can be negative. The fact that support is defined in terms only of the relation between the conditional probability  $P(h/e)$  and the prior  $P(h)$  implies that the so-called 'dynamic assumption', that after receipt of  $e$  (and no stronger

information), your degree of belief in  $h$  should be  $P(h/e)$ , is unnecessary, at any rate from the point of view of the Bayesian account of inductive inference.

The only endogenously determined quantities in the Personalistic Bayesian theory are the likelihood terms  $P(e/h)$  in the conditions stated above, and the probabilities of necessary truths and falsehoods relative to the individual's background information. The prior probabilities in Bayes's theorem are not usually of this type and are, therefore, parameters undetermined within the theory. This fact often leads people to believe that the account Personalist Bayesians give of inductive inference is subject to no significant constraints at all, and is consequently irremediably 'subjective', a mere branch of psychology, and not very good psychology at that. I shall simply ask the reader to wait until the final section for the discussion—and rebuttal—of this well-worn accusation.

### 3. The Falsity of the Null-Support Thesis

Since arguments have been put forward for the null-support thesis based on what appear to be unobjectionable methodological precepts, we perhaps ought first to see where these break down. There seem to be two such arguments: (i), advanced by Giere and Zahar, is that if  $h$  was designed to explain, and at the very least to be consistent with  $e$ , then  $h$  stood no chance of being refuted by those facts described by  $e$ . Hence, it is concluded,  $h$  cannot be supported by  $e$ , since support by  $e$  allegedly implies the possibility of refutation by  $e$ . The second argument, (ii), is due to Giere and Redhead, who assert that when the data are used as an explicit constraint in the process of constructing a hypothesis, then the probability of the data given the falsity of the hypothesis cannot be small, but on the contrary must be unity. But evidence, they contend, supports a hypothesis that predicts it when and only when it would be improbable if the hypothesis were false.

This criterion embodies a quite fundamental intuition. It is the intuition behind significance testing, for example: the null hypothesis (i.e., the negation of the proposed causal hypothesis, usually identified merely with the hypothesis that the observed data are due 'to chance') is rejected, and the causal hypothesis correspondingly confirmed, if those data are very improbable on the assumption that the null hypothesis is true; and—for future reference—it follows immediately from Bayes's theorem, in the form

$$P(h/e) = \frac{P(h)}{P(h) + \frac{P(e/\sim h)P(\sim h)}{P(e/h)}},$$

as entailing a high posterior probability of  $h$  when, as is assumed in Giere's and Redhead's discussion,  $e$  is probable given  $h$ . In general, though, as the expression

above makes clear, it is merely the likelihood ratio  $P(e/\sim h):P(e/h)$  that has to be small to confer a high posterior probability on  $h$ .

I shall not quarrel with either the improbability criterion, or the principle enounced in (i), that support for a hypothesis  $h$  from an experiment (a term I shall use to describe any data source) requires the possibility that the experiment is capable in principle of generating also refuting outcomes. Indeed, I shall be claiming that both these are naturally explained by the Bayesian theory. What I quarrel with is the claim that from either or both follows the null-support thesis, and I shall show that that claim is false.

Let us start with (i). Giere states that

if the known facts were used in constructing the model and were thus built into the resulting hypothesis . . . then the fit between these facts and the hypothesis provides no evidence that the hypothesis is true [since] these facts had no chance of refuting the hypothesis. (1984, 161)

Glymour (1980, 114) voices a substantially identical opinion, and much the same occurs, in a slightly more elaborate way, in Zahar (1983, 245) (incidentally, ignore the fact that statistical theories are not strictly refutable; we can take these authors to be using the word 'refute' in a sense that accommodates weaker criteria than the purely deductive). Plausible though it may sound, the argument is quite specious. How can *facts* ever have a chance of refuting anything? If  $e$  is a factual statement and  $h$  a hypothesis, then it is simply false to say that  $e$  has a chance of refuting  $h$ ; it either refutes  $h$  or it doesn't, and it does so or doesn't whether  $h$  was designed to explain  $e$  or not. Giere has confused what is in effect a random variable (the experimental setup or data source  $E$  together with its set of distinct possible outcomes) with one of its values (the outcome  $e$ ). It is only  $E$ , not  $e$ , that has the chance of refuting any particular hypothesis. Moreover, it makes perfectly good sense to say that  $E$  might well have produced an outcome other than the one,  $e$ , it did as a matter of fact produce. It follows that whether or not  $h$  was deliberately designed to explain  $e$  but nevertheless does so,  $E$  (in general) *could*, on the occasion on which it produced  $e$ , have generated another outcome inconsistent with  $h$ . Obviously, once  $E$  is performed and  $e$  results, there is no chance of that performance of  $E$  refuting  $h$ ; but, as we have seen, this would be true whether  $h$  was designed to explain  $e$  or not. Either way, (i) collapses.

(ii) fares no better. A propos Mendel's use of the observed ratio of tall to dwarf pea plants in the second filial generation of his famous experiment, Giere remarks that "fitting this case was . . . a necessary requirement for any model to be seriously entertained. So there seems no way [in which the data could be improbable given the negation of Mendel's factorial hypothesis]" (1984, 118). Let us not dispute Giere's rather doubtful assumption that Mendel invented his theory to account for the data (though too exact agreement with those data caused Fisher (1936), in a famous paper, to conclude that the theory was in fact constructed

first). But we should certainly dispute the validity of his argument, for the conclusion is a non sequitur. I at any rate can see no reason, and Giere provides none, why the data should not have been regarded as improbable on the supposition that Mendel's hypothesis was false. It might of course be argued that as the data are by assumption already known to have occurred, their probability must of necessity be equal to one, and hence equal to one conditional on the negation of Mendel's hypothesis. Glymour (1980), in a well-known argument, does indeed charge the Bayesian with having to accept just this conclusion.

I shall discuss Glymour's claim later, and argue that it is false. However, even were it true, it would not be a good strategy for supporters of the null-support thesis to adopt, since it would mean that there could be no way of distinguishing the support of hypotheses by data already known at the time both hypotheses were proposed, and that one hypothesis was designed to explain and the other explains independently. In either case, the probability of the data relative to the negation of both the hypotheses would be one, and neither would be supported according to the criterion—contrary to the declared opinion of virtually all the advocates of the thesis.

Writing, however, from an allegedly Bayesian position, Redhead (1986) seems to provide the linking argument Giere needs, but does so only by transforming Giere's premise, that constructing a hypothesis to explain *e* is tantamount to making the explanation of *e* a necessary condition for any hypothesis to be seriously entertained, into the explicitly Bayesian, and very strong, condition that in such cases *e* acts as a 'filter' allowing only those hypotheses nonzero prior probability which, modulo a set *a* of auxiliary hypotheses and other statements asserting that suitable initial conditions have been satisfied, entail *e*. From this considerably strengthened form of Giere's premise we do indeed infer that

$$P(e/\sim h \& a) = 1.$$

(I have followed Redhead here in writing *a* explicitly in the form of a condition in the probabilities), since for any partition  $[h_i]$ ,

$$P(e/a) = \sum P(e/h_i \& a)P(h_i/a) = \sum P(h_i/a) = 1,$$

where the filter condition ensures that the only  $h_i$  contributing to the sum is such that  $P(e/h_i \& a) = 1$ . It more or less immediately follows that  $P(e/\sim h \& a) = 1$ . Hence it does seem, when *h* is constructed in order to explain *e*, that the condition that *e* be improbable relative to *a* and the denial of *h* is never satisfied.

But it only seems that way. First of all, Redhead's filter condition is, as it stands, an impossibly strong condition, for it assigns a tautology zero prior probability (a tautology does not imply *e*). Even if tautologies are excluded by fiat, then if *h* implies *e* one can always find a nontautologous consequence of *h* that doesn't, and that must apparently then be assigned a zero probability where *h* is assigned a positive one, contradicting the probability calculus also. Presumably, the filter

condition is intended only to apply to the hypotheses in some partition. Whichever partition is chosen, however, it is easy to see that the filter condition still has the consequence that  $P(e/a) = 1$ . But this implies that  $P(h'/e \& a) = P(h'/a)$ , where  $h'$  is any hypothesis that entails  $e$  modulo  $a$ , whether  $h'$  was constructed to do so or not, so that  $h'$  cannot be confirmed by  $e$  either. In other words, Redhead's filter condition, even relativized to an 'appropriate' partition, is still too strong, for it cannot make the discrimination between the two types of hypothesis he wishes to make.

To sum up: nobody has made out a tenable case for supposing that when  $h$  is constructed to explain  $e$ ,  $e$  cannot therefore be regarded as improbable on the supposition that  $h$  is false. On the contrary, there seems no reason at all why this condition should not be satisfied. Take the Mendel case, for example. Mendel observed a fairly exact and stable ratio of tall to dwarf peas, whose occurrence in just those conditions corresponding to his careful selection of and mating the parent and first generation plants, for which only his theory of inherited, independently, and equiprobably segregated factors seemed to offer an explanation. In other words, the probability of obtaining such data, were Mendel's account not the correct one, should certainly not be unity. Of course, Mendel's theory was not widely regarded at that time as receiving great support from the data. But the explanation is not in  $P(e/\sim h)$  being unity, but in the contemporary implausibility of the particulate model that contradicted the favored blending theory.

In fact, the null-support thesis explains nothing, for it is false. Counterexamples abound, and we do not even have to go to the history of science to find them: they can be invented *ad lib*. The following two, one statistical and the other deterministic, should suffice. They are extremely simple in structure (one almost laughably so); and this is an advantage from more than the purely expository point of view, for it means that their salient characteristic can be diagnosed immediately.

(a) An urn contains an unknown number of black and white tickets, where the proportion  $p$  of black tickets is also unknown. The data consists simply in a report of the relative frequency  $r/k$  of black tickets in a large number  $k$  of draws with replacement from the urn. In the light of the data we propose the hypothesis that  $p = (r/k) + \epsilon$  for some suitable  $\epsilon$  depending on  $k$ . This hypothesis is, according to standard statistical lore, very well supported by the data from which it is clearly constructed. It is worth briefly going into the reasons why we take it to be well supported. They are that we are employing as a background theory the hypothesis that the 'experiment' has the structure of a sequence of so-called Bernoulli trials, in which  $p$  is the binomial parameter, which is to say that the draws are assumed to be independent with constant probability  $p$  of getting a black ticket.  $r/k$  determines a confidence interval of length  $2\epsilon$  for  $p$ , where the confidence level together with  $k$  determines  $\epsilon$ . I am not particularly concerned here with the ultimate epistemic rightness or wrongness of regarding these intervals as actually justify-



ing confidence of the relevant degree. My concern is simply with what is actually and uniformly regarded as legitimate practice, and there is no question but that confidence interval estimates of physical parameters, derived via some background theory involving assumptions about the form of the error distribution, are the empirical bedrock upon which practically all quantitative science is built.

But we can point to a feature of the hypothesis we have derived about  $p$ , which is, I submit, highly germane to an explanation of its epistemic merit. This is that the probability of the sample data relative to the same background distribution, but on the assumption that the parameter  $p$  lies somewhere outside the specified interval, is very much smaller than its small probability on the assumption that  $p$  does in fact lie in that interval. Recall that this is just the condition, endorsed by the Bayesian theory, for a high posterior probability for that interval to contain  $p$ . Let us leave the discussion in abeyance and now look at (b).

(b) The urn remains the same, but instead of sampling with replacement we now sample without replacement, and continue until the urn is empty. The proportion  $p$  of black tickets we discover to be  $p_0$ , and this now becomes our hypothesis about the value of  $p$ . Surely in this case the sample data support, since together with background information to the effect that the urn has remained the same throughout, they entail the hypothesis that  $p = p_0$  (and a fortiori the probability of the sample data on the assumption that that hypothesis is false is zero, given the background information).

While it is difficult to maintain that (b) is representative of much of quotidian scientific inference, both it and the more representative (a) are nevertheless very instructive. They are both cases where background theory supplies a model of the experiment that leaves only a parameter to be calculated from the data, and that background theory is sufficiently firmly entrenched to be taken more or less for granted (though additional data may conceivably lead to its being questioned, nonetheless). Continuity considerations would therefore seem to suggest that in general, the support of a hypothesis  $h(a_0)$  obtained from a parametric hypothesis  $h$ , whose adjustable parameter(s) is (are) evaluated as  $a_0$  from the data, should depend on the prior plausibility of the parametric model  $h$ .

I shall argue that this is indeed the case, with the help of an example that might be thought a rather surprising choice. In this rather simple and idealized example, the data from which  $a_0$  was evaluated are going to be data that are quite uninformative about the truth or falsity of the model  $h$  itself. So: let  $h$  be the hypothesis that two observable variables  $x$  and  $y$  are related linearly, so that  $y = cx + d$ , for some  $c, d$  that are to be determined by observation; and suppose also that we have no reason to believe that, within an interval determined by background information, any one set of values is any more likely than any other. Let  $e$  consist of two independent joint observations of  $x$  and  $y$ . Thus  $e$  determines, up to some interval

depending on the error distribution over the observations, values of  $c$  and  $d$ , which we shall represent by  $a_0$ .

Computing the degree to which  $e$  supports  $h(a_0)$  as the difference  $e$  makes to your assessment of the likelihood of  $h$  (not using that term in its specialized statistical sense), it is easy to see, even without assuming that these 'quantities' are represented by numbers as opposed to the members of some arbitrary additive semigroup, that this support may be considerable, though it is bounded above by the prior credibility of the model  $h$ . For the credibility of  $h(a_0)$  in the light of  $e$  is, given the assumption of the mutual independence of  $h$  and  $e$ , no more and no less than the prior credibility of  $h$ , since all  $e$  then does is evaluate the parameters  $c$  and  $d$ . But the credibility of  $h(a_0)$  independently of  $e$  is negligible, since  $(c, d)$  can, we may assume, take any values in the plane (usually there will be some prior restriction on their possible values, but these may well be very broad indeed). It follows that were we to possess exactly the same information as we do now, with the exception of a knowledge of  $e$  itself, then the adjunction of  $e$  would usually make some, and possibly a considerable, difference to our evaluation of the credibility of  $h(a_0)$ . In other words, the potential of  $e$  to alter the credibility of  $h(a_0)$  in an otherwise identical knowledge situation will vary with the prior plausibility of  $h$ .

But surely this proves too much – for how could  $e$  possibly support  $h(a_0)$  when  $e$  by hypothesis provides no information relevant to the truth of  $h$ ? The answer simply is that  $e$  *does* support  $h(a_0)$  to the extent that it raises its probability in general. I suspect that the apparent force of the objection derives from covertly assuming the truth of the so-called Consequence Condition, that if evidence supports a hypothesis then it must support every logical consequence of that hypothesis. I have in effect just presented an argument that I believe shows the Consequence Condition to be false. Indeed, as Popper and Miller have shown (1983), if we make increase in probability the criterion for support, then every hypothesis supported by some data has a logical consequence that is actually countersupported, in the sense that its probability is decreased, by that data. The Consequence Condition is in conflict also with more basic intuitions. We do believe, I think, that the approximate constancy of the acceleration induced in falling apples supported Newton's gravitational theory, but also that it did not support the bare hypothesis that the gravitational force is not everywhere constant.

It is, I believe, just because support for  $h(a_0)$  tends to be conflated with support for  $h$  that the null-support thesis is so firmly entrenched: what seems to happen is that the null support for  $h$  in cases like the above gets illicitly transferred to  $h(a_0)$ . The conflation is apparent, for example, in John Worrall's grounding his conclusion that "of the empirically accepted logical consequences of a theory those, and only those, used in the construction of the theory fail to count in its support," on the alleged fact that "Mercury's perihelion [advance] is not regarded as supporting classical theory," although it is predicted by versions of that theory

(1978, 48). Classical theory may or may not be supported; but even were it not supported, it certainly would not follow, as Worrall's inference presupposes, that the versions of classical theory that predict the perihelion advance are not supported by it.

Now let us observe that the informal, intuitive reasoning above is perfectly mirrored in the Bayesian theory. The assumption of the mutual irrelevance of  $h$  and  $e$  translates into the condition of probabilistic independence:  $P(h \& e) = P(h)P(e)$ . Note also that, modulo initial conditions,  $h(a_0) \langle = \rangle h \& e$ ; and substituting appropriately into Bayes's theorem we obtain  $P(h(a_0)/e) - P(h(a_0)) = P(h)[1 - P(e)]$  (Howson 1984, 248–49). Thus  $P(h)$  is an upper bound on the support, which is positive so long as  $P(e) < 1$  (we assume that  $P(h) > 0$ ).

The dependence of the support of  $h(a_0)$  by  $e$  on the prior probability of  $h$  (since the support depends on the prior probability of  $h(a_0)$ , which implies  $h$ ) is quite general in the Bayesian theory, and is reflected in the judgments of working scientists. The statistician and biometrician Karl Pearson discovered a family of density curves (his Type I, II, III, IV, and V curves), which he was prone to fit to a great variety of data, in a way that to many of his contemporaries seemed frankly ad hoc: on one such occasion the economist Edgeworth pointedly asked "what weight should be attached to this correspondence by one who does not perceive any theoretical reason for those formulas?" (Edgeworth 1895). Kepler fitted ellipses to Tycho's data on planetary orbits, but he also thought it necessary to present independent reasons for that type of orbit. Nearer to home we find Kitcher (1985) castigating a piece of sociobiologists' parameter adjustment on the ground that "the model gives absolutely no insight into the reasons behind the periodicity [the adjusted parameter] . . . the choice of a periodic function for the probability bears no relation to any psychological mechanisms" (375). And so on; anyone can find a host of examples.

The application to the initial problem of discriminating the support of the ad hoc variants (1) from that of  $y = f(x)$  is now clear. The introduction of the parameters  $a_1, \dots, a_n$ , which are subsequently evaluated from the data, corresponds to postulating models

$$y = f(x) + (x - a_1) \dots (x - a_n)g(x)$$

of the experimental situation that we simply have no reason to believe true – and because the set of these 'models' is infinite, and for none of them is there the remotest reason to believe it true; the credibility of each is literally zero. Thus the situation with the hypotheses (1) falls under the general case of evaluating the support of  $h(a_0)$ , where the prior credibility of the model  $h$  is zero, or effectively so. This conclusion may seem a little disappointing: the hypotheses (1) are not supported by the data, because in effect we don't think the sort of structure they postulate is the true one. Isn't an assessment of support supposed to justify rather than be justified by our convictions as to what is likely to be true and what isn't?

We have seen many attempts to construct theories of confirmation that claim a priori status. But nothing comes out of nothing; and all these theories incorporate just such convictions, often, as in Carnap's systems, in a disguised form, as basic principles. Jeffreys's theory (expounded in his [1948]) delivers the judgment that the parametric hypotheses of which (1) are instances are a priori less likely to be true than  $y = f(x)$ . This judgment is an immediate consequence of his Simplicity Postulate, which asserts as a general principle that of the hypotheses advanced within science, those with fewer undetermined parameters (simpler in Jeffreys's sense) are a priori more likely to be true than those with more. However, Jeffreys himself did not see the Simplicity Postulate as an a priori valid principle; for him it was rather an explicit recognition of scientific practice: the simplest equations are as a matter of fact preferred when fitting curves to data (1948, 10).

In the unlikely event of absolutely no background information about the experimental source, simplicity by itself may well play a role in determining levels of support among the uncountably many possible functional relationships consistent with the data. Clearly, if a proposed curve fits some initial set  $e_1$  of observations, and continues to fit all subsequent sets, we should want to say that in so doing it becomes increasingly well supported; but this is possible only if its probability prior to all the observations were positive. Since the curve is likely to have been relatively simple among all the possibilities, we are in such circumstances implicitly taking simplicity to be a ground for assigning a moderate or at any rate nonzero prior probability. But in general, where there is a body of background information constraining the plausible candidates, simplicity and prior probability may well not march in step. So the Simplicity Postulate must be rejected as a principle of general scope. (Its rejection has often been urged on the grounds of alleged inconsistency, most recently by Watkins [1985, 110–16]; that charge, I have argued elsewhere [1988], is incorrect.)

Popper's well-known reversal of Jeffreys's probability ordering must also be rejected. Popper's reason for adopting the converse ordering is that being more easily tested, simpler hypotheses are less probable than more complex ones. But this is to confuse pragmatics with epistemology: we simply have no ground a priori for believing that more easily testable hypotheses are less—or for that matter more—likely to be true, and we should certainly not allow strong and, in principle, ungroundable epistemological assumptions, which these in fact are, to play the role of logical axioms.

Questions of ultimate justification are, however, beside the point of this exercise, which is the much more limited one of diagnosing the differential status we accord the hypotheses (1) compared with the initial  $y=f(x)$ ; and we have shown that the fact that the former were generated using the data and the former is not in itself the cause. The crucial feature of the variants (1), which accounts for their comparatively low status, is that the *introduction* of the parameters  $a_1, \dots, a_n$

has no justification in terms of what we think likely to be true (the same point is made in Nickles [1985, 200; and 1987]).

#### 4. Prediction and Accommodation; the Bayesian Analysis

I have argued that, depending on circumstances, hypotheses can be and often are regarded as supported by data employed as constraints in their construction. The circumstances can be summarized in the condition that  $P(e/\sim h)/P(e/h)$  be small. This condition is, of course, as I pointed out earlier, just the Bayesian condition for a high posterior probability of  $h$ .

The null-support thesis is false. The motivation for it was a desire to disqualify certain types of patently ad hoc accommodation by a theory of otherwise adverse, or at best neutral, data. A doctrine weaker than the null-support thesis, but similarly motivated, concedes that while accommodated data may give some support, it is nevertheless never to the same extent as if the data in question had been independently predicted. This is also false, and again it is not difficult to manufacture informal counterexamples to it. Consider the following (in essence due to Peter Urbach). A numerologist employs a number of algorithms, in a manner claimed to represent the vagaries of divine will, to predict dates of major earthquakes in California, where 'major' means exceeding some given Richter value. This goes on year after year, failures of the phenomenon to occur to order being explained away suitably, until eventually one such prediction comes true. Established geophysical theory predicts (let us suppose; the moral does not depend on factual accuracy) that earthquakes of such magnitude occur when and only when the strain along a fault line exceeds by some specified quantity a critical value. There is no independent way of estimating when this value will be exceeded. I think that there are few people who would credit the numerologist's theory with greater support from the observed phenomenon than the assertion that on the date at which that phenomenon was observed the strain exceeded the critical value by at least the amount specified by standard theory. And the reason why we do not regard the independently predicting hypothesis as having no support here is the same as the reason why we regard the data-constructed hypotheses (1) as having no support in the circumstances in which they arose: we simply don't believe that they can be true. Being constructed or not from the data has nothing to do with it.

Although the general thesis that an independently predicting hypothesis is always better supported by the data so predicted than is one that is deliberately constructed to explain them is false, there is an important residuum that is not. Because it lends itself to a perspicuous treatment, let us again consider the example of three hypotheses  $h'$ ,  $h$ , and  $h(a_0)$ .  $h$  as before contains an undetermined parameter that is evaluated from  $e$ .  $h$  is inconsistent with  $h'$ , which independently predicts  $e$ ;  $h(a_0)$  only 'predicts'  $e$  after the event ( $a_0$  is the parameter in  $h$  evaluated from  $e$ ). Finally, we shall suppose that the prior probabilities of  $h$  and  $h'$  are

equal. In more idiomatic language,  $h$  and  $h'$  are rival explanatory frameworks;  $h'$  predicts the effect  $e$ , and so does  $h(a_0)$ , but only as a consequence of  $e$ 's having antecedently been used to calculate a free parameter in  $h$ :  $h'$  predicted, while  $h(a_0)$  is merely an accommodation of, the data. In these circumstances it seems correct to say that  $h'$  picks up more support from  $e$  than does  $h(a_0)$ .

This is—or at any rate seems to be—the residuum of truth in the false general thesis that independent prediction invariably scores higher in terms of support than accommodation. It is certainly regarded as true from the point of view of the Bayesian theory; indeed, it is very easily generated as a consequence of that theory, where support  $S$  is measured as simply the difference of posterior and prior probabilities. For where the initial conditions are regarded as being part of background knowledge, we have  $S(h'/e) = P(h')[1 - P(e)]/P(e)$ , and  $S(h(a_0), e) = P(h(a_0))[1 - P(e)]/P(e)$ . But by assumption  $P(h') = P(h) \geq P(h(a_0))$  and so  $S(h', e) \geq S(h(a_0), e)$ . Admittedly, the inequality is weak, and so to say that  $h'$  picks up strictly more support from  $e$  is strictly incorrect. Never mind; the result is good enough.

The sufficient condition for the inequality must not be forgotten; it is that  $h$  has at most the prior probability of  $h'$ . If for example the parameter in  $h$  is introduced purely ad hoc to yield the desired effect, this ad hocness will be registered in the prior probability of  $h$  being small if not negligible (such would have presumably been the case with the de Sitter modification of Poincaré's Lorentz-invariant gravitational theory, which contained a parameter specifically introduced to explain the annual shift in Mercury's perihelion). It must be emphasized that this is merely a special case, though I suspect that the apparent plausibility of the thesis that independent prediction always gleans more support than accommodation rests on nothing more than invalidly generalizing from it.

## 5. The Objections to Personalistic Bayesianism

The Bayesian theory seems to offer a most promising formal reconstruction of our intuitive reasoning in the contexts we have discussed. It is, furthermore, to my knowledge the only methodological theory that is capable of making sense of our intuitions, to say nothing of canonical practice, in those, and other, contexts. But objections, and on the face of it powerful ones, have been brought against its credentials to perform such a reconstructive role.

One such objection is highly relevant to the problem we started with, of assessing the supportive power of known facts relative to theories whose construction is carried out with those facts employed as explicit constraints. The objection, which I alluded to earlier, in section 2, is that if  $e$  is already known then both  $P(e)$  and  $P(e/h)$  ought to be set equal to one, and not only then do all the Bayesian formulas we have written down adopt trivial forms, but in particular the support (2)

simply goes to zero, so that the support of hypotheses by known data, whether they were designed to satisfy the data or not, is uniformly zero.

The objection originates with Glymour (1980, Chap. 3). It is, I believe, based on a misunderstanding of how the Bayesian formulas are intended to be interpreted, and it has a straightforward and natural answer (and one, we shall see, that Glymour himself anticipates). This answer is as follows. The Bayesian claims that the support that  $e$  gives a hypothesis  $h$  is to be evaluated by the extent to which the observation of  $e$  alters the prior credibility you attach to  $h$ . What does this mean when  $e$  is already known? It cannot mean that that the  $P(e)$  and  $P(e/h)$  terms are trivially one, since  $e$  is *always* known at the time you compute the support function: you would be guilty of a simple misapplication of the theory, rather analogous to dividing both sides of an equation by zero, if you *therefore* made those probabilities unity. What  $P(e)$  is intended to convey, whether  $e$  is known or not, is how likely you think  $e$  *would* be were (i)  $h$  true and (ii)  $h$  false; and  $P(e)$  is simply a weighted sum of these two magnitudes.

To the extent, then, that support is a function of  $P(e)$ , it is a function not of an actual probability, but of a subjunctively characterized one; and when  $e$  is known, the subjunctive conditional characterizing it becomes counterfactual: how probable do you think  $e$  would be if you didn't already know it to be the case relative to the suppositions, respectively, that (i)  $h$  is true, and (ii)  $h$  is false? There is absolutely nothing ad hoc or in conflict with core Bayesian principles in defining support in this way. On the contrary, it seems a very natural way of proceeding. Certainly the presence of subjunctives in the definition of the constituent probabilities is nothing new: the definition of conditional probabilities is also cast in the subjunctive mood, as we observed earlier.

Glymour himself considers this response to his objection quite sympathetically, but he is doubtful as to whether adequate consistent procedures exist for computing these probabilities. He considers various methods for calculating such values and concludes that they do not work. I am quite willing to concede that it is difficult if not impossible to come up with a sharp value for the probability I would attribute to a stone's falling to the ground when dropped if current gravitation theory were false, but this is just one of those occasions when the deliverances of the Bayesian rule are going to be very imprecise. Moreover, the existence or otherwise of algorithms or general criteria of evaluation is beside the point for the sort of Bayesianism I am considering here, which is not a source of rules for computing all the probabilities figuring in Bayes's theorem, but simply an attempted reconstruction of a type of intuitive reasoning, in which more or less rough estimates of various probabilities are made. Some of these can be analyzed as involving the application of a particular rule, and others cannot, and have to be treated simply as exogenously determined data, among which typically are the prior probabilities  $P(e)$  and  $P(h)$ .

The relegation of  $P(e)$  and  $P(h)$  to the status of exogenous parameters whose

values in many cases seem not to be susceptible of any even moderately precise determination might well seem to invite the charge of triviality. Take the mere presence of the undetermined parameters  $P(e)$  and  $P(h)$  first. These have frequently been identified as a source of weakness in the Personalistic Bayesian theory, allegedly undermining any claim it might make either to explanatory status (those parameters can be adjusted to ensure consistency with practically any historical judgment of what has supported what: due to their presence the theory is allegedly no more than "a soft and rubberlike system which is easy to manipulate") or to objectivity (they can be adjusted in a way that, e.g., "allows [people] to assign zero probability to a promising rival hypothesis that threatens ones they personally favour"). (Both quotations are from Watkins [1985, 308].)

Let us address these objections in turn. First, the presence of undetermined parameters does not preclude a hypothesis's either being tested or having the capacity to explain phenomena. Every scientific theory of note has some parameters undetermined within the theory. It has to be conceded that it is often not possible to arrive at anything like precise measurements of individuals' prior probabilities; but this does not distinguish that account from other quite respectable explanatory theories, where sometimes the estimated values of parameters amount to nothing more than a well-grounded qualitative assumption. How many times, for example, do we see explanations in the physical sciences and elsewhere prefixed with remarks like "the masses may be considered to be so small that the potential energy of interaction is zero," or "suppose  $a \ll b$ ; then . . ."; and so forth. Qualitative assessments of belief for which there is independent evidence can support just as good explanations in the Bayesian account.

But what about the charge of extreme subjectivism? It is, after all, the incapable subjectivity of Bayesian assessments of support, depending as they do on the individual's priors, that alarms people most, and that inspired Fisher's famous and influential verdict that those assessments are measures "of merely psychological tendencies, theorem concerning which are useless for scientific purposes" (Fisher 1947, 6–7), echoed more recently in Jaynes's verdict that "personalistic probability belongs to the field of psychology and has no place in applied statistics" (Jaynes 1968, 231). These dismal conclusions are quite unwarranted, however. Partly they result from a simple non sequitur. The subject matter of the Personalist Bayesian theory is beliefs, and beliefs are, of course, psychological. But the theory is not psychology: it sets out to describe not beliefs as such, but the structure of beliefs regulated by consistency constraints that are completely objective. The Personalist Bayesian theory is, in other words, a *logic* of beliefs; it consists in setting out conditions of consistency that are no less objective than those of deductive logic. To be fair, even the defenders of the Personalistic theory do not emphasise—or even appear to recognize—its unimpeachably logical status. They almost invariably concede the charge of outright subjectivism, and try to mitigate it by appeal to the very general phenomenon of asymptotic conver-



gence of the posterior probabilities relative to the same data. The charge should not, I stress, be conceded in the first place.

But it must be conceded that judgments of support do, according to the Personalistic theory, reflect to a greater or lesser extent the influence of one's own prior belief distribution. Before one condemns this as amounting to a betrayal of objective standards, one should ponder the status of principles that affect to determine substantive inductive judgments. These principles, however 'objective' they purport to be, are inevitably assumptions, and moreover somebody's assumptions, and honesty compels that they should be presented as such. The Personalistic theory merely calls a spade a spade. If anybody still doubts that it is necessary to give explicit recognition to the role of undefended prior belief in inductive inference, they should examine with care (as Peter Urbach and I do [1988]) those theories, like classical statistics, for example, that claim to dispense with it. I suspect that they will eventually, though possibly reluctantly, recognize that those theories do not deliver the goods. They should also recall that they do not dismiss deductive logic because it refrains from supplying criteria for justifying premises as well as inferences. Deductive logic contents itself with judgments of the form: "If you wish to remain consistent, then if you believe this set of statements to be true, you must also accept that statement as true." The same acceptance of the limitations of the power of human reason implicit in the restriction of scope here is unfortunately still not evident in 'objective' discussions of inductive inference.

Let us now return to the ability of the Bayesian theory to explain characteristic modes of inductive inference. It is raised in an apparently acute form by some empirical studies of subjects' evaluations of statistical data, and in a well-known survey Kahneman and Tversky express a strongly negative conclusion:

The usefulness of the normative Bayesian approach to the analysis and the modeling of subjective probability depends primarily not on the accuracy of the subjective estimates, but rather on whether the model captures the essential determinants of the judgment process . . . In his evaluation of evidence [however] man is apparently . . . not Bayesian at all. (Kahneman and Tversky 1982, 46)

I cannot go into the detailed evidence that is taken to support this conclusion; it would take far too long. I shall simply concede that there are areas in which popular modes of reasoning fail to satisfy the Bayesian constraints (so much, incidentally, for the claim that the theory imposes none). So what? Statistical inference is an area in which popular reasoning is notoriously subject to simple fallacies. Wason's celebrated card paradox (Wason 1966) exhibits an area where popular deductive intuitions also fail in just such uniform ways. Kahneman and Tversky might wish to conclude from this that people are not deductive logicians either; and to an extent they would be right. But that is not a conclusion that should disturb those who claim that the canons of deductive logic exercise visible

constraints on the way people reason, especially after some reflection, in a great variety of circumstances. We know that they recognize, at least in principle, that the characteristic of deductive inferences that makes them valuable is that they preserve truth, and it would be strange indeed if they were unable to exercise this knowledge in cases that make not too great demands on their powers of reasoning.

The same ought to be true of probabilistic reasoning, and the criteria of consistency that apply there. The mathematical theory of probability, as a matter of historical fact, started life as the theory of fair odds, and as such was immediately applied to the problem of inductive inference by the seventeenth- and eighteenth-century mathematicians. The fact that the criteria developed by the theory are not uniformly applied by everybody does not mean that people in general are not capable of recognizing the authority of those criteria, nor that in many simple cases they are incapable of applying them.

I am going to consider one final objection to this exercise in explanatory Bayesianism (a more complete discussion of all these, together with other objections, is to be found in Howson and Urbach [1988]). This is not to say that there are not others, but the list is a long one, and a criterion of importance has to be exercised if this paper is not to get too tail heavy. I shall not, therefore, discuss the status of Jeffreys-conditionalization, or other proposed modifications of the 'dynamic assumption', since they do not really conflict with any principle I have invoked here, nor shall I discuss David Miller's (1966) charge, now known as Miller's paradox, that the way the Bayesian evaluates likelihoods in terms of the values of a given physical probability distribution is inconsistent. (Graham Oddie and I have, I think, shown in Howson and Oddie [1979] that Miller's reasoning rests on a standard type of fallacy.)

This final objection is one with which readers of the earlier discussion will now be familiar. It is that to whatever extent we may invoke estimates of credibility in our inductive reasoning, we do so qualitatively, and it is beyond question that we could never honestly refine these to points in the real continuum. Yet the Bayesian theory is a theory of point-valued subjective probabilities. It has been proposed (e.g., by Koopman [1940], Good [1962], Smith [1961], Dempster [1968], and Williams [1976]), in response to such observations, that a theory of interval-valued probabilities would furnish a more realistic foundation for a theory of subjective uncertainty, and these authors have indeed developed such a theory, which now tends to go under the name of the theory of upper and lower probabilities (the upper and lower probabilities are the end points of the intervals). The theory is a generalization of the point-valued theory; the latter is obtained when upper and lower probabilities are identical. It is not clear, however, that more realism is imported into such a theory, since the upper and lower probabilities themselves are point valued, and it seems no easier in principle to arrive at nonarbitrary values for these than for point probabilities themselves.

Should we attempt to find more realistic weakenings of the point-probability

model? Personally, I do not think so. Certainly the objection that we cannot realistically claim to make point-valued probability assessments should not force us to abandon the point model. For that objection is not at all as strong as it sounds; it is, on the contrary, very frail. In the first place, there is nothing in the Personalist theory that says that anyone does make point estimates: that theory is quite compatible with people's actual evaluations being as crude and as qualitative as you like. Moreover, the people who take this objection seriously commit themselves thereby to a position on the use of real number theory in empirical science that they ought, on reflection, to find wholly untenable. For if the fact that point values are in principle incapable of being arrived at is taken to invalidate such explanatory claims, then we should have to give up obedience to the laws of real arithmetic as the explanation of why we employ our customary modes of calculation where any *physical* magnitude is concerned. Although we are quite happy with the usual mathematical theory of length, volume, etc., that makes these quantities real-valued, it is nevertheless a fact that the length of, e.g., a room, is not, even discounting the practical impossibility of arbitrarily precise measurement, a real number, nor even an exact nondegenerate interval of real numbers. Our estimates of such magnitudes are made only to within a nondegenerate interval possessing no exact upper and lower limits, which *in principle* cannot be refined to a single point. This does not invalidate a theory of these magnitudes that postulates real-number values, or render meaningless calculations based on real arithmetic; we should not, therefore, feel obliged either to regard the theory of point-valued probabilities as invalidated by similar facts about the nature of subjective probability.

## 6. Conclusion

Methodological theories wax and wane, like the scientific hypotheses that are their subject matter. The mathematical treatment of uncertainty based on the probability calculus was one such; announced in the pages of the Port Royal Logic and some writings of Leibniz, it rapidly developed in scope and sophistication, and reached its apogee in the early years of the nineteenth century. It enabled you to compute the odds in games of chance, the odds on the sun rising the next day, and the odds on coincident testimonies of witnesses being the result of their telling the truth. By the early twentieth century the theory was virtually dead, the victim of untoward facts – not empirical facts, to be sure, but logical ones. It was inconsistent, and the inconsistency was, so it appeared, at the very heart of the theory. Yet half a century later it is back again, risen like Lazarus, and fast recruiting disillusioned members of more recent faiths.

Amended to give the theory now known as Personalist Bayesianism, it is demonstrably consistent. The cost of consistency was abandoning the famous method, called by Keynes the "Principle of Indifference" and by von Kries the

"Principle of Insufficient Reason," for obtaining ostensibly 'informationless' prior probabilities. The Principle of Indifference prescribes uniform probability or probability-density distributions over bounded parameter-spaces as the mathematical representation of ignorance, but it became apparent that these distributions are far from invariant under different ways of representing the hypothesis space, with the result that the same hypothesis may be assigned different values depending on the space in which it is embedded. Worse, one and the same hypothesis may be assigned different values under logically (or, to be more precise, logico-mathematically) *equivalent* representations of the hypothesis space.

People have tried to bring back suitably watered-down versions of the principle (they are known as Objective Bayesians). I have argued in the foregoing sections, however, that no loss in desirable strength results from simply declining to prescribe any criteria for determining prior probabilities. The addition of such criteria merely burdens the theory with indefensible assumptions, even where they are consistent. Without them, I have argued, we possess a logic of confirmation that admits that our theories are no more than theories, and seeks to discover what can be constructed according to the sole criterion of consistency. And that, I have tried to show, is quite a lot. Had Popper not shown such animosity toward the enterprise of bringing subjective probabilities into epistemology, and had he not laid down a priori his own stultifying criterion—a criterion that I have argued at length (Howson 1973, 1987) is just as indefensible as the Principle of Indifference that it much resembles—for prior probabilities, that in all interesting cases they should be zero, he might therefore have recognized in Personalist Bayesianism a genuine, purely deductive logic of confirmation, yielding nontrivial information, yet free of the sort of synthetic inductive principles he rightly declared should have no place in methodology.

### References

- Dempster, A. P. 1968. A Generalisation of Bayesian Inference. *Journal of the Royal Statistical Society B*, 30: 205–32.
- Edgeworth, F. Y. 1895. On Some Recent Contributions to the Theory of Statistics. *Journal of the Royal Statistical Society* 58: 505–15.
- Edwards, W. 1968. "Conservatism in Human Information Processing." In *Formal Representation of Human Judgement*, ed. Kleinmuntz. New York: John Wiley.
- Fisher, R. A. 1936. Has Mendel's Work Been Rediscovered? *Annals of Science* 1: 115–37.
- . 1947. *The Design of Experiments*, 4th ed. Edinburgh: Oliver and Boyd.
- Giere, R. N. 1984. *Understanding Scientific Reasoning*, 2d ed. New York: Holt, Rinehart and Winston.
- Glymour, C. 1980. *Theory and Evidence*. Princeton: Princeton University Press.
- Good, I. J. 1962. "Probability as the Measure of a Non-Measurable Set." In *Logic, Methodology and Philosophy of Science*, eds. Nagel, Suppes, and Tarski. Stanford: Stanford University Press.
- Howson, C. 1984. Bayesianism and Support by Novel Facts. *British Journal for the Philosophy of Science* 35: 245–51.

- . 1973. Must the Logical Probability of Laws be Zero? *British Journal for the Philosophy of Science* 24: 133–163.
- . 1987. Popper, Prior Probabilities and Inductive Inference. *British Journal for the Philosophy of Science* 38: 207–24.
- . 1988. The Consistency of Jeffreys's Simplicity Postulate, and its Role in Bayesian Inference. *Philosophical Quarterly* 38: 68–83.
- Howson, C., and Oddie, G. 1979. Miller's So-Called Paradox of Information. *British Journal for the Philosophy of Science* 30: 253–78.
- Howson, C., and Urbach, P. M. 1988. *Scientific Reasoning: The Bayesian Approach*. LaSalle, Ill.: Open Court.
- Jaynes, E. T. 1968. Prior Probabilities. *Institute of Electrical and Electronic Engineers Transactions on Systems Science and Cybernetics* SSC-4: 227–41.
- Jeffreys, H. 1948. *Theory of Probability*, 2d. ed. Cambridge: Cambridge University Press.
- Kahneman, D., and Tversky, A. 1972. Subjective Probability: A Judgement of Representativeness. *Cognitive Psychology* 3:430–54.
- Koopman, B. O. 1940. The Bases of Probability. *Bulletin of the American Mathematical Society* 46: 763–74.
- Kitcher, P. 1985. *Vaulting Ambition*. Cambridge: MIT Press.
- Miller, D. W. 1966. A Paradox of Information. *British Journal for the Philosophy of Science* 17: 59–61.
- Nickles, T. 1985. Beyond Divorce: Current Status of the Discovery Debate. *Philosophy of Science* 52: 117–207.
- . 1987. Lakatosian Heuristics and Epistemic Support. *British Journal for the Philosophy of Science* 38: 181–205.
- Popper, K. R., and Miller, D. W. 1983. A Proof of the Impossibility of Inductive Probability *Nature* 302: 687f.
- Redhead, M. L. G. 1986. Novelty and Confirmation. *British Journal for the Philosophy of Science* 37: 115–18.
- Smith, C. A. B. 1961. Consistency in Statistical Inference. *Journal of the Royal Statistical Society B*, 23: 1–25.
- Wason, P. 1966. Reasoning. In *New Horizons in Psychology*, ed. B. M. Foss. Harmondsworth, U.K.: Penguin.
- Watkins, J. 1985. *Science and Scepticism*. Princeton: Princeton University Press.
- Williams, P. M. 1976. Indeterminate Probabilities. *Formal Methods in the Methodology of Empirical Sciences*, eds. Przlecki, Szaniawski, and Wojcicki. Dordrecht: Ossolineum and Reidel.
- Worrall, J. 1978. "The Ways in Which the Methodology of Scientific Research Programmes Improve on Popper's Methodology." In *Progress and Rationality in Science*, eds. Radnitzky and Anderson. Dordrecht: Reidel.
- Zahar, E. G. 1983. Logic of Discovery or Psychology of Invention? *British Journal for the Philosophy of Science* 34: 243–61.
- . 1989. *Einstein's Revolution*. LaSalle, Ill.: Open Court.